Supplementary Materials

Here, we provide some descriptive analysis of the Flickr traces: the distribution of trip distances showing a power-law with exponential cutoff behavior, some visualizations of the largest Flickr flows in both the air-travel and commuting networks, compared against the largest flows in the ground-truth datasets. Then, we evaluate the performance of the gravity model and the hybrid gravity model in terms of the Sørensen-Dice coefficient.

Finally, we visualize the performance of the radiation model and the hybrid radiation model by showing their prediction ratio as a function of distance, the scatter plot of the predicted values against the ground-truth values and their performance under data availability constraints. Such figures complement those in the main text for the gravity model and the hybrid gravity model.

Distribution of Flickr users' trip distances

The trip distances follow a power-law distribution with exponential cut-off, as shown in Figure S1. This distribution is in accordance with previous research on individual human mobility *(1)*.
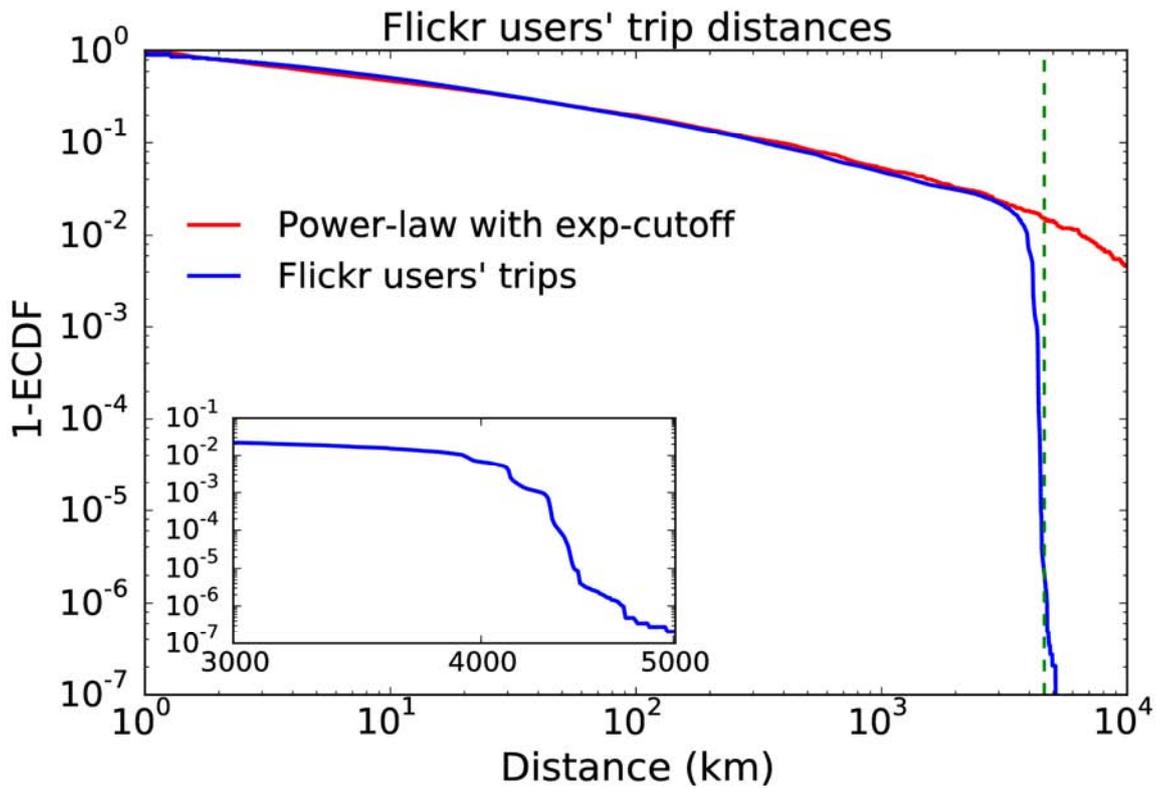
Fig. S1. Trip distance distribution. The trip distances of Flickr users follow a power-law distribution with exponential cut-off, bounded by the extension of the continental U.S., whose diameter (maximum distance between two points) is about 4,000 kilometers.

Large flows between airports predicted by Flickr

Figure S2 shows that the Flickr trips are highly correlated with the air travel dataset when observing the pairs of airport basins with large flows of passengers. The green flows represent the connections predicted to have more than 10,000 passengers by the Flickr model.
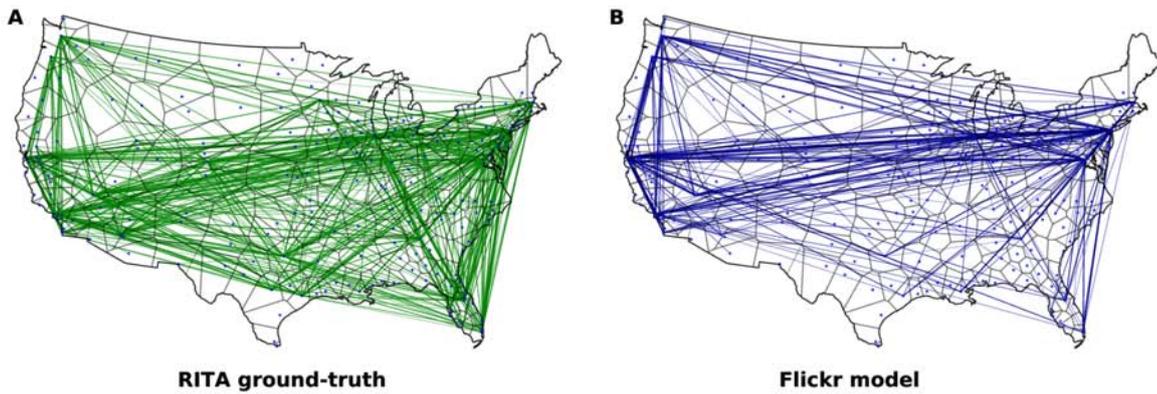
Fig. S2. Large air transportation flows captured by Flickr. (A) Air flight connections with more than 10,000 in the RITA dataset. **(B)** Flows with more than 10,000 trips as predicted by the Flickr model. Visualization created with matplotlib *(2)*.

Large commuting flows predicted by Flickr

Figure S3 shows that the Flickr trips are highly correlated with the commuting census data when observing the pairs of counties with large flows of commuters. The green flows represent the connections predicted to have more than 1,000 passengers by the Flickr model.
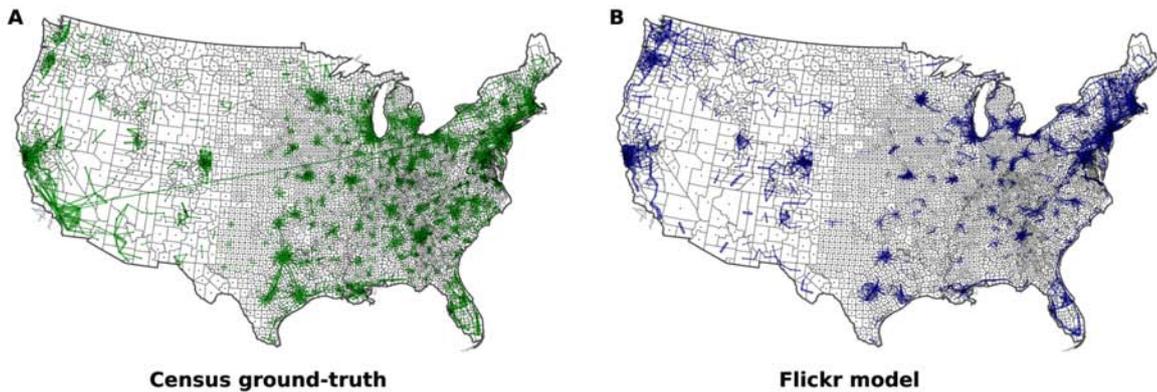
Fig. S3. Large commuting flows captured by Flickr. (A) Commuting flows with more than 1,000 in the U.S. Census dataset. **(B)** Flows with more than 1,000 trips predicted by the Flickr model. Visualization created with matplotlib *(2)*.

Distance thresholds

Fig. S4 compares the ground-truth flows for air travel and commuting against the Flickr flows at the county level and the airport level respectively. It shows that the Flickr flows have good agreement with the ground-truth for distances above 500km for air travel, and below 100km for commuting (panels **A**, **D**). In fact, if those flows are filtered from the data, then the Flickr users trip distance distributions become consistent with the ground-truth trip distributions, maximizing the Pearson correlation between both (panels **B**, **E**). The threshold distances are also revealed as the value for which the correlation between the real flows and the Flickr predicted flows is maximum (panels **C**, **F**).
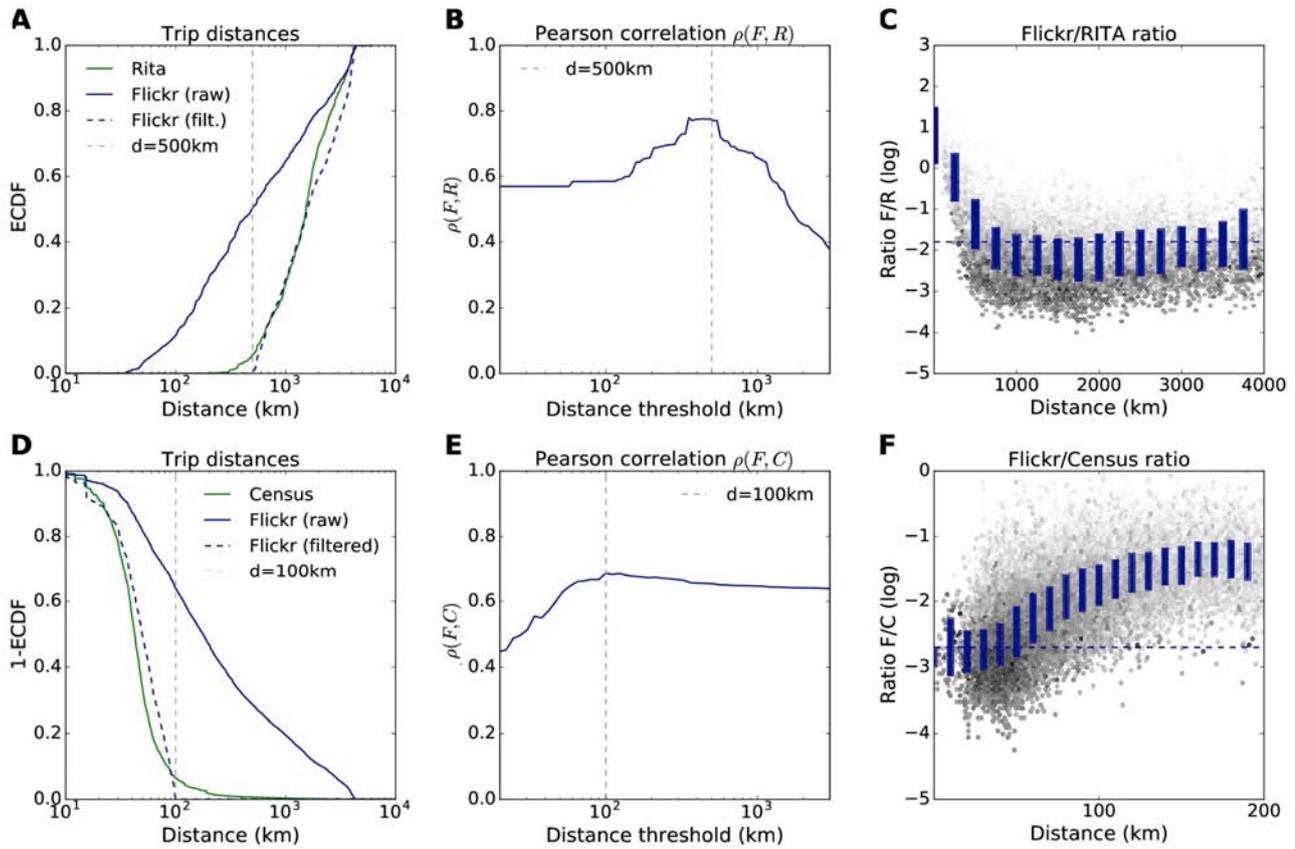
Fig. S4. Distance calibration for the U.S. air transportation network and the U.S. commuting network. A correct fitting of the human mobility networks based on geolocalized traces requires aggregating the latter at the appropriate resolution level (basins or counties, respectively) and filtering those ones associated with the correspondent mobility type. Distance is an important calibrator for this task, as can be seen from **(A, D)** the cumulative distribution of the trip distances in Flickr compared with the ground-truth (the RITA dataset in the case of air travel, and the census information for commuting); **(B, E)** the Pearson correlation between Flickr flows and the ground-truth flows for trips above a distance threshold; **(C, F)** the ratio between Flickr and the ground-truth flows for a pair of (source, destination) nodes --basins or counties-- as a funtion of the distance between them. Here, dashed lines represent the linear regression coefficient between Flickr and the ground-truth flows. **(A, B, C)** Air transportation network; **(D, E, F)** Commuting network.

Spatial cross-validation of the model: U.S. West Coast vs. U.S. East Coast

To extend the analysis of the generalizability of the model, we performed a geographical 2-fold cross-validation in which we split the contiguous United States into two roughly symmetric parts, taking by reference the meridian -102°. We trained the model in one half of the U.S. and we then used it to predict the mobility flows inside the other half. To make our test independent from the specific choice of the spatial partition, we disregarded crossed flows, that is, the traffic flows connecting two points in different halves of the partition. Table S1 shows the performance in terms of the Pearson correlation ρ and the determination coefficient r^2. We see that, in this case, the performance of Flickr traces is inferior with respect to the 10-fold cross-validation values of Table 1, denoting the presence of spatial heterogeneities in the users' activities across the U.S., especially for the commuting network. However, the performance of the hybrid gravity model is still improved by the assimilation of the Flickr flows.

Model	Commuting		Air travel	
	ρ	r^2	ρ	r^2
Gravity model	0.72	0.41	0.68	0.40
Flickr model	0.60	0.34	0.72	0.43
Hybrid gravity model	0.78	0.46	0.81	0.49

Table S1. Geographically cross-validated model performance (U.S. West Coast vs. U.S. East Coast). Performance of the hybrid gravity model in terms of the Pearson correlation coefficient ρ and the determination coefficient r^2, when the flows are predicted by training the model in one of the Coasts and validating it in the other.

A Sørensen-Dice coefficient based test

In the main text we thoroughly explored the performance of the hybrid model under geographical and random sampling *(bootstrapping)* constraints. Here we analyze the performance for different flows subsets organized by distance and destination population, using a test based on the Sørensen-Dice similarity coefficient. The Sørensen-Dice similarity between two sets *A* and *B* is defined as:

$$s(A,B) = \frac{2|A \cap B|}{|A| + |B|}. \text{ (S1)}$$

Here we follow the modification introduced in *(3)* as named as *"Common part of commuters"* (*CPC*) and also used in *(4, 5, 6)* in order to compare the predicted flows $\mathbf{H} = (h_{ij})$ against the real flows $\mathbf{Y} = (y_{ij})$. Following Eq. 6 in *(4)*:

$$CPC(\mathbf{H},\mathbf{Y}) = \frac{2\sum_{ij} \min(h_{ij}, y_{ij})}{\sum_{ij} h_{ij} + \sum_{ij} y_{ij}}. \text{ (S2)}$$

Fig. S5 shows the goodness of fit in a grid of *(distance, population)* ranges, where each cell represents a subset of pairs of origin-destination basin filtered by distance and by destination population. In the left and central pictures, darker green colors represent a higher similarity between the predicted and real flows; the right pictures show in red those cells that were improved by the assimilation of Flickr traces, while those in blue are worse fitted when using Flickr data. We observe that in the air travel network (lower pictures) the improvement of the hybrid model is almost constant, but it is more evident for large population basins. In the commuting network, the gravity model outperforms the hybrid gravity model for small cities (up to ~10,000 inhabitants), but for larger cities the hybrid gravity model gives better predictions. However, by improving the largest flows of people (which are usually associated to highly populated cities) the hybrid gravity model can give a better estimation of the total number of human movements at both resolution levels.
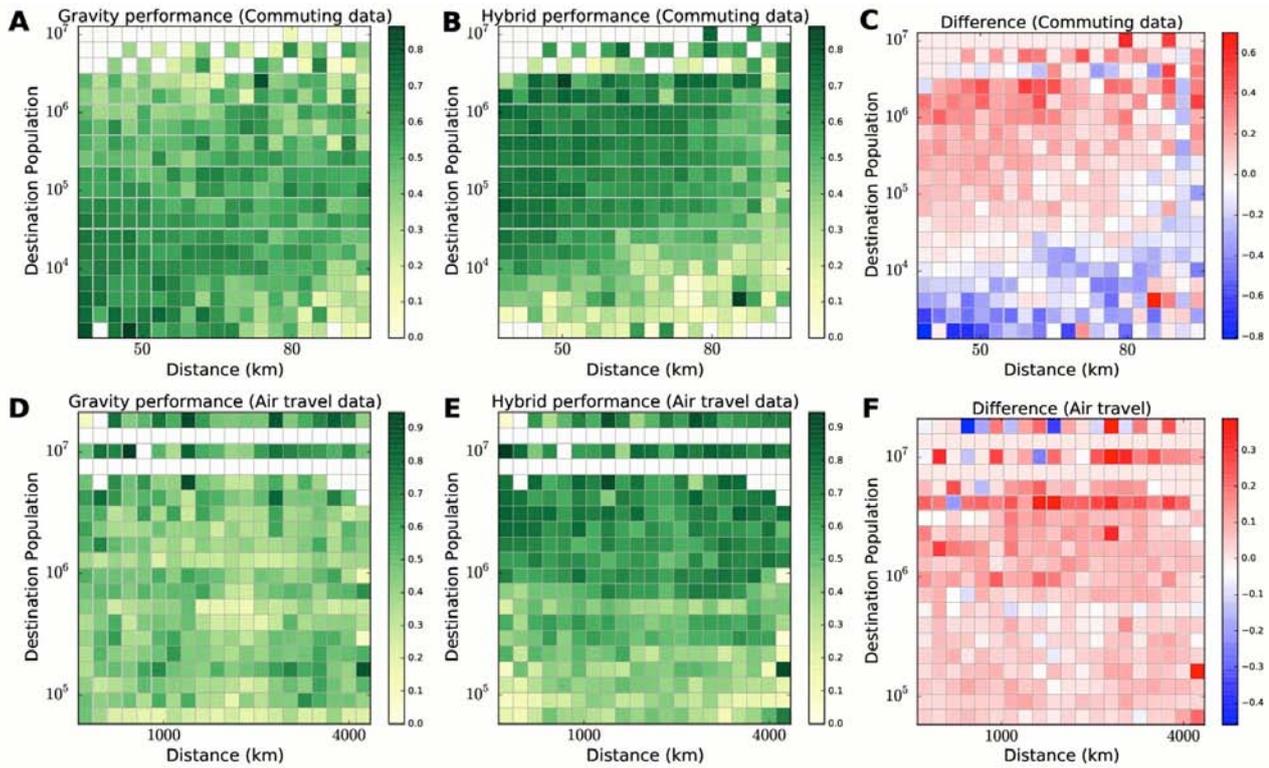
Fig. S5. Sørensen-Dice coefficient grids for the gravity model and the hybrid gravity model. Flow prediction error measured by the Sørensen-Dice coefficient between pairs of nodes under different *(distance, destination population)* cells. Prediction errors are computed for: **(A, D)** the gravity model; **(B, E)** the hybrid gravity model. Darker green represents a higher similarity and thus a smaller error. Panels **(C, F)** represent the difference between both models. **(A, B, C)** Commuting network. **(D, E, F)** Air travel network.

Performance of the hybrid radiation model

The radiation model was used as input for a stacked regression similar to the one proposed for the gravity model in the main text (see Equation 1). The assimilation of Flickr traces produced a hybrid radiation model which was trained and 10-fold cross-validated using the ground-truth data. In Figure S6 we show the prediction ratio for the radiation model, the Flickr model and the hybrid radiation model. The radiation model shows a good behavior for the commuting network, but it strongly underestimates large distance flows in the air transportation network, as the number of intervening opportunities is dominated by other factors such as city relevance and population size of the destination. The assimilation of Flickr traces is not enough to solve the issue in the air travel network, and the gravity model should be preferred in this case. Figure S7 shows the scatter plots of the radiation model and the hybrid radiation model for both networks.

Finally, Figure S8 shows the performance of the hybrid radiation models in terms of the determination coefficient and in the prediction of large flows. The radiation model curves are horizontal because the model is parameter free, so it does not undergo a training process in the stacked regression. For the air travel network the radiation model alone has a negative determination coefficient which is not shown in the plot. For the commuting network, we observe that a 3% of Flickr traces is already beneficial for improving the performance in the hybrid radiation model, confirming what we observed for the hybrid gravity model.
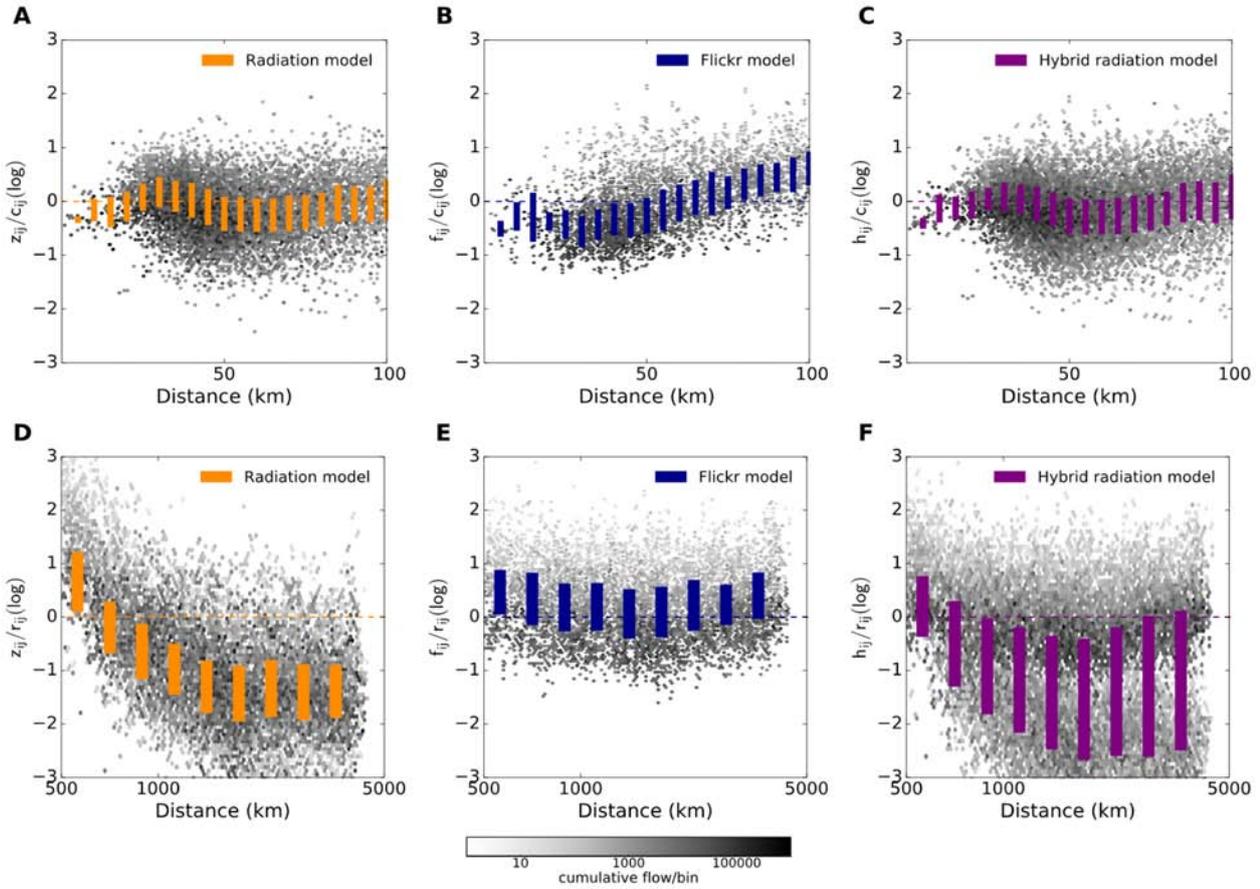
Fig. S6. Prediction ratios for the radiation and hybrid radiation models. In each panel, grey dots correspond to the flows between basins as predicted by the radiation model (z_{ij}), by the Flickr model (f_{ij}), and by the hybrid radiation model combining both (h_{ij}), in relation to the real flows, and as a function of distance between basins. Flows predictions were all made under a 10-fold cross-validation scheme except for the radiation model alone, which does not require a fit to the data. **(A) (B) (C)** Commuting network of the U.S. **(D) (E) (F)** Air-transportation network of the U.S. Only flows above 100 passengers are shown, and the color intensity at each point represents the total passenger flows aggregated under a certain distance and prediction ratio. Boxplots represent interquartile ranges.

10
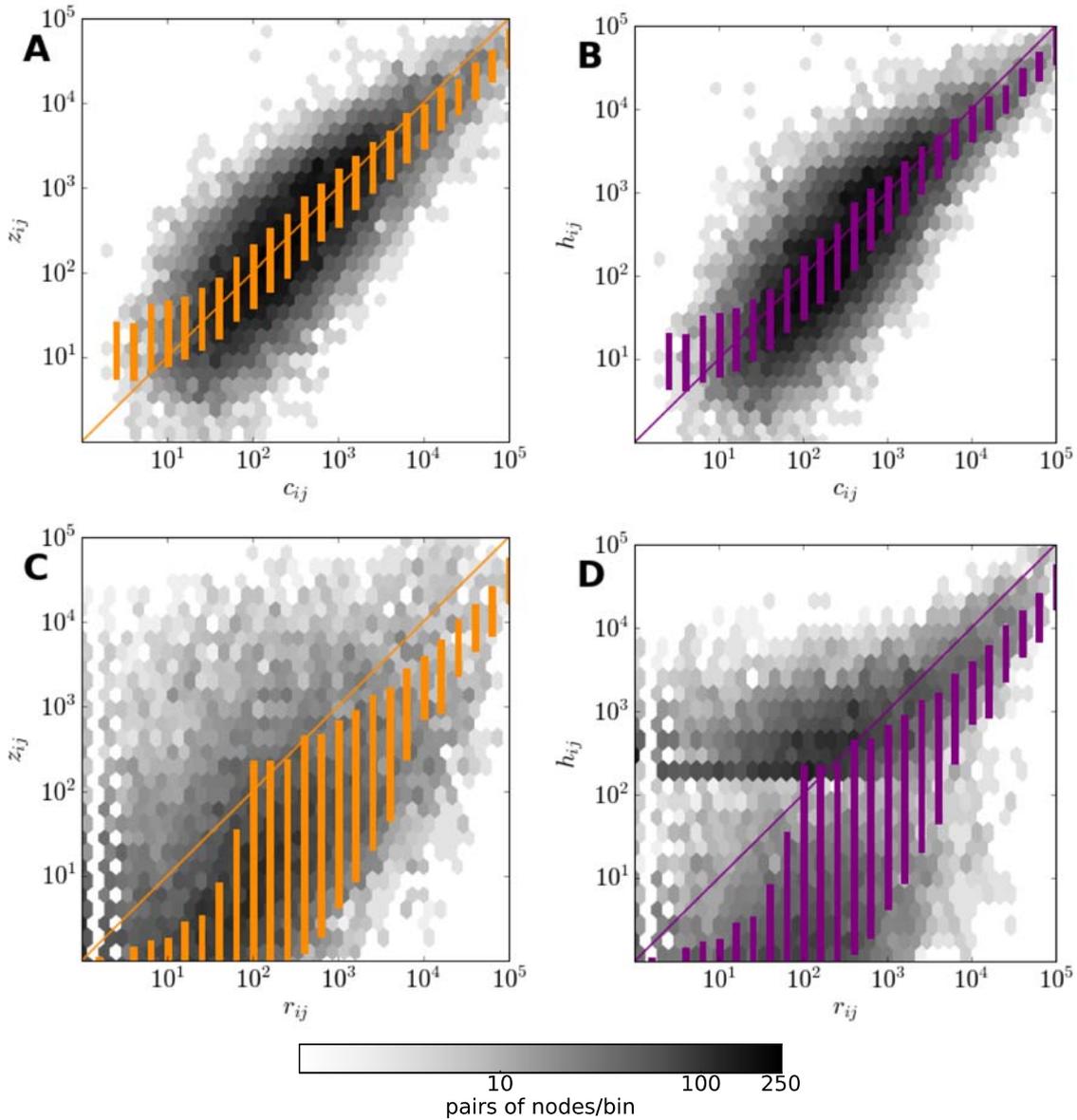
Fig. S7. Radiation models predictions. The 2D-histograms compare the predicted values against the real flow values for **(A, C)** the radiation model; **(B, D)** the hybrid radiation model. Each point represents flows with some real/estimated flow value relation. Points color in a gray scale represents the frequency values. The boxplots correspond to the interquartile range within a bin. **(A, B)** U.S. Commuting network. **(C, D)** U.S. air travel network.
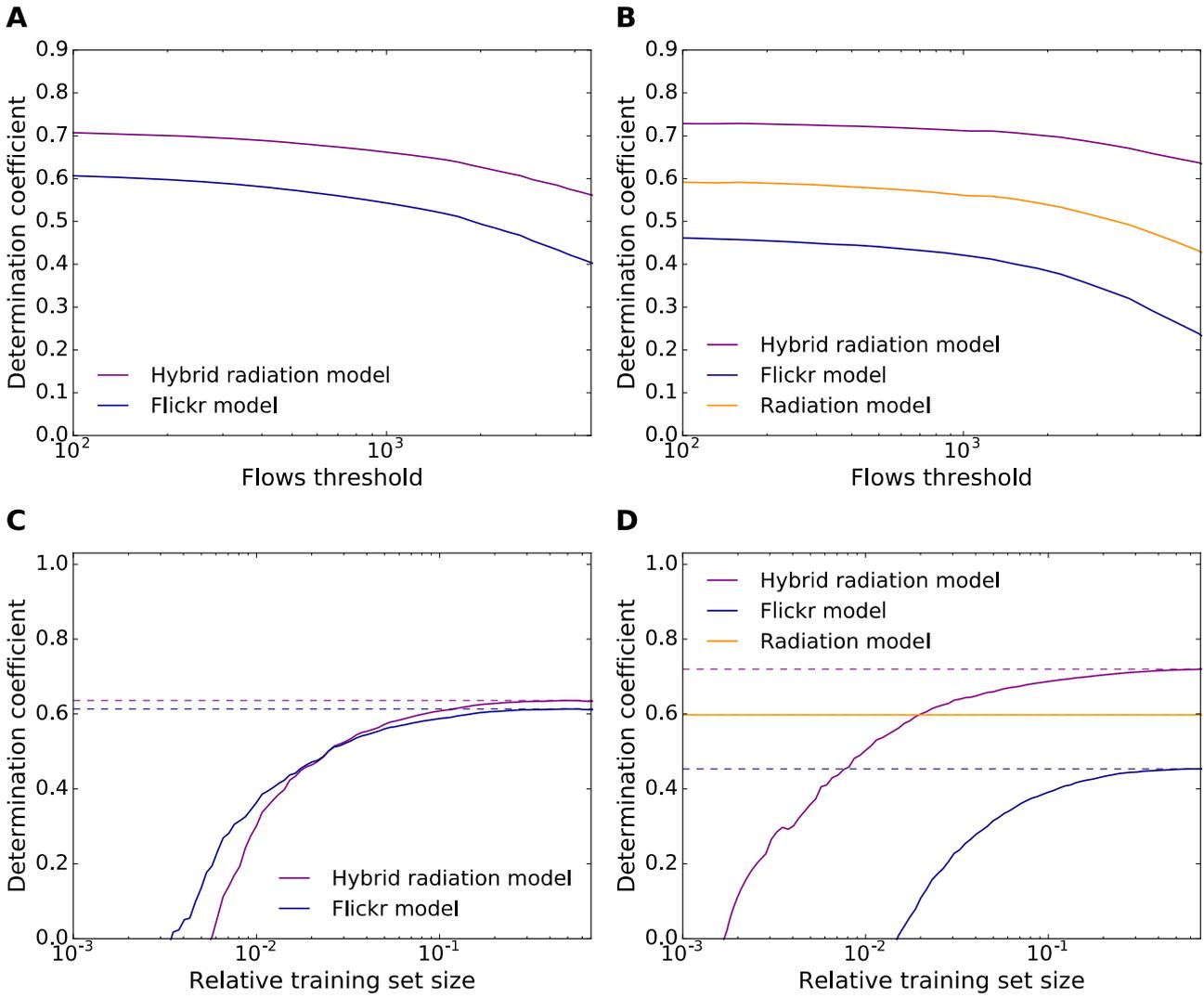
Fig. S8. (A, B) Determination coefficient for the prediction of large flows, as a function of the flow threshold. Performance of the models measured by the determination coefficient r^2 when restricted to pairs of nodes for which the real flows are above a given threshold value. **(A)** Air transportation data. **(B)** Commuting data. **(C,D) Determination coefficient as a function of the training set relative size.** The training set sizes were varied in a logarithmic scale. Dotted lines correspond to the maximum attained values. **(C)** Air transportation data. **(D)** Commuting data.

References

1. Gonzalez MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779-782

2. Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9(3):90-95

3. Simini F, Maritan A, Néda Z (2013) Human mobility in a continuum approach. PLoS ONE 8(3):e60069

4. Masucci AP, Serras J, Johansson A, Batty M (2013) Gravity versus radiation models: on the importance of scale and heterogeneity in commuting flows. Phys Rev E 88:022812

5. Yang Y, Herrera C, Eagle N, González MC (2014) Limits of predictability in commuting flows in the absence of data for calibration. Sci Rep 4:5662

6. Wesolowski A, O'Meara WP, Eagle N, Tatem AJ, Buckee CO (2015) Evaluating spatial interaction models for regional mobility in Sub-Saharan Africa. PLoS Comput Biol 11(7):e1004267